

ACI Proactive Risk Manager®  
Enterprise Fraud Management Solution

2019 Data Science Professional Services Offerings

Data Science Technology and Process Overview

Table of Contents

**Table of Contents**

[Overview 3](#_Toc17289077)

[Model Development 3](#_Toc17289078)

[ACI Data Science Team 3](#_Toc17289079)

[Business Requirements 3](#_Toc17289080)

[Automation and Pipelines 4](#_Toc17289081)

[Data Validation 4](#_Toc17289082)

[Risk Table Generation 4](#_Toc17289083)

[Feature Generation 5](#_Toc17289084)

[Feature Analysis and Transformation 5](#_Toc17289085)

[Iterative Model Optimization 6](#_Toc17289086)

[Production Model Delivery 6](#_Toc17289087)

[Data Science Roadmap 7](#_Toc17289088)

[Roadmap Trajectory 7](#_Toc17289089)

[Continuous Learning 8](#_Toc17289090)

# Overview

Enterprise class machine learning models represent one of the core fraud detection components in ACI Proactive Risk Manager®. The process ACI employs to develop these models is the result of over 23 calendar years and 80 staff-years of machine learning algorithm, analytics process and software design and development, evolving and growing as a result of dozens of large-institution model development and deployment projects.  Over time, each project has contributed to the expansion of the suite of model development features, utilities, strategies, processes and model deployment methodologies, culminating in the process our team of data scientists employs today.

ACI implements two types of machine learning models for Proactive Risk Manager customers: custom models and segment models.

* Custom models – Custom models are developed on the full dataset of the financial institution.  Custom models benefit from a global view of the customer’s genuine and fraud transacting patterns, important to achieving higher predictive performance since fraudsters operate across all FI products and channels.  We optimize our custom models to detect specific segments of fraud, segments which achieve the overall business and operational requirements for financial institution.
* Segment models – These models are developed on a specific segment, or data strata, where possible segmentation criteria include PIN versus signature, international versus domestic, debit versus deposit fraud and others. ACI offers segment models to individual financial institutions for whom there is a business case, to enable that FI to achieve higher fraud detection rates at lower false positive rates.

Following is an overview of the model development process. Each primary stage will differ for a custom model versus a segment model and will be customized depending upon the customer’s business requirements and channel-specific data.

# Model Development

## ACI Data Science Team

ACI has a dedicated team of M.S and Ph.D. level data scientists who are specialists in card present and card not present model research and development. The team consists of highly experienced principal data scientists, as well as senior and junior data scientists. Multiple staff have strong theoretical academic and research backgrounds, necessary for the ongoing applied data science research projects which ACI progresses, some of which are summarized in this document. Collectively this team has over 80 staff-years of experience applying the use of machine learning algorithms within the payments, banking and retail fraud detection sectors. The Data Sciences team is augmented and supported by ACI’s risk analysts and by senior software engineers in the Application Development team.

## Business Requirements

Each project starts with identification of the customer’s risk management and fraud prevention business requirements, operational requirements which define every major stage of the model development process, including within-sample and out-of-time performance quantification and expected production performance. Examples of business requirements include target product and portfolio, primary predictive performance metrics, fraud score operating point, target alert volume, real-time configuration, treatment of declines and others. All requirements are reflected in the Business Requirements Document and jointly agreed with the customer and ACI.

## Automation and Pipelines

To achieve the ongoing roadmap objective of decreased model delivery time, while simultaneously implementing improvements for higher model predictive performance, ACI has automated a significant portion of what was previously a manual data preprocessing and model optimization process. ACI employs pipelines to achieve this automation, and has replaced a number of the manual processes previously completed by the human expert with machine learning algorithms. Machine learning algorithms are implemented as features, or inputs to the machine learning model, and are also used in a number of the feature preprocessing and model optimization stages of the process, resulting in significant reductions in model development time while simultaneously supporting higher model predictive performance. A subset of the processes in the descriptions which follow are implemented with these algorithms.

## Data Validation

After all data is transferred to ACI, a senior data analyst validates the full dataset of the financial institution and prepares it for model development. As data quality and preparation strongly influences model quality, ACI applies the required labeling, sampling and data transformation to ensure the highest quality data is used to develop the model. Label quality is critical to the outcome of the process, and ACI employs fraud data available in the Proactive Risk Manager FraudAth table, TC40/SAFEand PIN fraud, when available. Data validation and preparation consists of the following:

* Labeling – ACI employs a custom algorithm to ensure discover and labeling of the most fraud auths while avoiding labeling genuine cards as fraud.
* Down-sampling genuine cardholder – Generation of multiple train, test and validation datasets
* Field and feature configuration – The senior data analyst acquires a deep understanding of all relevant data elements and prepares for configuration of the thousands of features to be included in the project
* Descriptive statistics – The senior data analyst generates a wide variety of descriptive statistics on the data, including analyses of outliers and missing values, a subset of which are included in the Data Validation Presentation.
* Data validation presentation – The primary customer-facing deliverable of this stage of the project is a presentation which describes the model development dataset in detail, with a focus on fraud behaviors.

## Risk Table Generation

Risk tables quantify the risk level of individual, combined or aggregated values of categorical fields, both across all data and within defined data segments. Risk is represented as a normalized ratio and can be based upon both incidence and amount. Segmented risk tables are generated on defined data segments, for example international and domestic, and provide the model with more granular risk measures for the target data element.

Some risk ratios are simple, normalized measures of risk, where others are generated by a separate class of machine learning model. The latter is able to capture and store more complex, probabilistic relationships between data elements which cannot be represented with the simpler form of risk tables. Each type of risk table capture differing types of genuine and fraud patterns, some across all data and other more granular, and are collectively complementary. Risk tables form the foundation of a broad class of features and are required for the next model development stage, feature generation.

## Feature Generation

In addition to data quality, features form the foundation of model predictive performance. Features are derived data variables, generated from the raw authorization and demographic data. Features capture fraudulent and genuine behaviors within the data, behaviors which enable the model to discriminate between the two target classes, fraud and genuine. ACI maintains a suite of thousands of features, all directly relevant to the fraud use case, and these features represent one of a number of core data science intellectual property to the organization.

ACI’s feature engine has evolved over the years, and represents one of a number of key data science investment areas, with multiple staff continually dedicated to enhancing the application. The feature engine has a number of benefits:

* Expressivity – The feature engine is highly expressive, capable of representing the spectrum of account behaviors, from the most basic to very complex. Examples of basic features include counting the frequency of specific types of banking / payment events within user-define time periods, and examples of complex features include features which aggregate and transform probabilistic risk ratios from risk tables described in the earlier section.
* Parameterization – Features can be categorized into more general themes, and ACI implemented these base feature themes within feature parameterization infrastructure. Feature parameterization enables the data scientist to instantiate thousands of features rapidly, for a wide variety of data types, cardholder behaviors, user-define time periods and risk tables. The benefit is not only a risk suite of features upon which to base a custom model, but reduced model time to market.
* Data transformation – It is sometimes necessary to transform data fields prior to or after generating the features which require those data elements. ACI’s feature engine enables a wide variety of data transformation, from simple data mapping to more complex conditional history reordering and filtering. This enables the engine to present the historical data to the feature suite for each account prior to feature generation, resulting in another implicit parameter for potentially each of the thousands of features generated.
* Custom features - A subset of projects include development of custom features, features which may benefit only one or a small number of customers. This mechanism enables the data scientist to develop features for a fraud or genuine pattern which they observe is specific to a dataset on which a custom model is being developed. This option is also necessary because ACI primarily provides custom models to our customers.

ACI’s feature engine provides additional benefits, which are represented in part in the Sample Model Predictive Performance section below.

## Feature Analysis and Transformation

After generating the thousands of features, the data scientist evaluates the features and generates a number of statistical measures on each. Two of the feature transformation steps which are implemented in this stage of the project include:

* Feature preprocessing – The data scientist preprocesses the features prior to executing dimensionality reduction, using all datasets include both missing data and outliers and extreme values, and these will be reflected in the feature values, a subset of feature ranges are modified prior to proceeding with the next stage of the project, dimensionality reduction. ACI implements this range optimization process with a separate machine learning model designed specifically for this purpose. As with other model development utilities, this range optimization utility executes as a multi-core process, decreasing executing time while still ensuring high quality feature ranges.
* Dimensionality reduction – The final model ACI delivers to each customer will normally consist of a small subset of features drawn from the full suite of thousands of features. The first step to determine that feature set is dimensionality reduction. ACI implements multiple machine learning algorithms to generate a collection of higher performing feature sets, feature sets which are input to the next stage of the project, iterative model optimization.

## Iterative Model Optimization

Having established a strong foundation in both the data preparation and validation, and then feature generation, analysis and transformation, the next primary stage of the project is iterative model optimization. Iterative model optimization represents another area of data science investment for ACI. ACI’s model optimization cluster enables the data scientist to develop thousands of candidate models on a dataset within approximately 5 days, resulting in the exploration of a greater portion of the feature combination / model parameter space.

To achieve this, ACI wrapped the core classification algorithms with a combination of machine learning algorithms and expert heuristics. The current infrastructure supports multiple, differing versions of wrapper algorithms which implement complementary model optimization and feature combination objectives. The data scientist determines within this stage of the project when to utilize each version of the optimization cluster, including where appropriate the application of manual tuning and feature pruning.

Following development of the hundreds or in some cases thousands of candidate models, the data scientist selects the highest performing model which achieves the business requirements defined at the start of the model development project.

## Production Model Delivery

Following model benchmarking, ACI’s Service Delivery team works with the customer in delivering the model, validating the installation and supporting the customer to repopulating the scoring engine database. The customer then transitions the model from the test to the production environment.

The primary project deliverables include:

* Business Requirements Document
* Data validation report
* Custom model and all required control file
* Model performance presentation detailing the performance of the final model according to a wide variety of industry metrics

# Data Science Roadmap

ACI maintains roadmaps for all solutions under the Universal Payments Framework, including PRM and all Data Science professional services as well as the core machine learning technologies implemented within the PRM Scoring Engine and Universal Scoring Engine. Primary data science roadmap objectives include:

* Data Science as a Service
* Higher model predictive performance
* Decreased model delivery time
* Agile delivery / CICD

Following is a summary of ACI’s overall roadmap trajectory as well as one highlight from the roadmap which will be used to achieve that trajectory, continuous learning.

ACI has not finalized the target delivery timeframe for these initiatives.

## Roadmap Trajectory

Regarding ACI’s Roadmap data science technologies, ACI is transitioning to a Data Science as a Service (DSAAS) fraud detection delivery model. DSAAS will provide financial institutions and online retailers with a simple, highly performant fraud detection service via continuous delivery of models and features without the need for the risk manager to deploy, manage, monitor and decommission individual fraud models. Using methods such as continuous learning, automated daily models and ensemble methods, the most recent fraud and genuine patterns are incorporated into the DSAAS score without the need for manual model development and deployment. DSAAS has the following benefits:

* A single, simple fraud detection mechanism generated by one or more models which do not require standard model management practices
* Models which generate the score are continually updated with the latest fraud and genuine patterns of the retailer with no human intervention
* Implements a continuous learning mechanism which updates all production deployed model parameters with confirmed fraud and genuine patterns as they occur
* All model retrains are fully automated, with no manual Data Scientist model retrain processes required
* All model characteristics, including the number of models, model segments, history of changes, features deployed and model parameters are transparent to the user via dashboards
* All model outputs are aggregated into a single score which consistently achieves all functional requirements, including consistent fraud detection rates, daily alert volumes, earliness of detection and transaction and card false positive rates
* Non-functional requirements, including throughput and latency, are validated for all model changes prior to automated production deployment
* Agile delivery including CICD (Continuous integration / continuous development)

## Continuous Learning

One component of ACI’s data science roadmap is continuous learning. Continuous learning is one of a number of advances required to achieve DSAAS, and represents an area of active research for ACI’s data science team. Continuous learning:

* Is a class of machine learning algorithms which allow for continuous modification of model parameters with new exemplars without the need for a full model train on a large, batch training dataset
* Is fully automated, not requiring data scientist execution of any stage of the model train pipeline, including production model deployment
* Is implemented within a continuous learning module / framework which includes model performance validation functions to ensure that only models having performance criteria defined by the user are deployed into production
* Provides full operational statistical stability, including stable daily transaction / card / merchant alert volumes, false positive rates and non-functional requirements
* Begins to address the threat of continuously changing fraud patterns and cybercriminal’s continued use of increasingly sophisticated attack vectors
* Also addresses the evolving nature of genuine patterns which, if not incorporated into a model, increase false positive rates
* Achieves all non-functional requirements, as well as ensures conformance to all retailer / financial institution Governance / Risk / Compliance (GRC) policies and processes

Continuous learning represents only one of a number of components in ACI’s data science roadmap. ACI would be glad to provide an overview of the full roadmap separately.



© Copyright ACI Worldwide, Inc. 2019

ACI, ACI Payment Systems, the ACI logo, ACI Universal Payments, UP, the UP logo and all ACI product names are trademarks or registered trademarks of ACI Worldwide, Inc., or one of its subsidiaries, in the United States, other countries or both. Other parties’ trademarks referenced are the property of their respective owners.